

<https://doi.org/10.7236/JIIBC.2022.22.5.147>
JIIBC 2022-5-22

인공지능 서비스 영향평가 추진방안에 대한 연구

A Study on Implementation Plan for AI Service Impact Assessment

신선영*

Sunyoung Shin*

요약 본 연구의 목적은 영향평가에 대한 정의부터 국내외 인공지능 서비스 영향평가 사례에 대한 분석을 바탕으로 인공지능 서비스 영향평가 추진에 대한 정책적 제언을 수립하는 것이다. 이를 위해서 국내외 타 분야에서 추진된 영향평가 사례, 인공지능 서비스 국내외 영향평가 사례를 바탕으로 추진 방향을 분석하였다. 국내 인공지능 서비스 영향평가는 다소 광범위하고 포괄적이며, 시점도 사전적 예방 수단에 그치지 않고 상시적·사후적 위험성 관리를 예정하고 있다. 단계별 추진 방안으로 1단계에는 AI 수준 조사 기반의 경제적 효과 등의 정량적 지표를 개발한 후, 2단계에서는 지능정보화 기본법에 기술된 안전성 및 신뢰성, 인공지능 윤리 등 정보문화, 고용·노동 등 사회·경제, 정보보호, 국민의 일상생활에 미치는 영향에 미치는 분야별 평가체계를 마련한다. 3단계에서는 세부 측정지표나 방식 등의 논의를 확대하고 영향평가 결과가 인공지능 정책에 반영하는 환류 체계 포함된다면 국내의 인공지능 경쟁력 강화에 도움이 되는 정책 수단이 될 수 있다는 것을 제시하였다. 본 연구는 향후 정책 설계자, 인공지능 서비스 개발자, 시민단체 등 다양한 참여자를 통한 분석이 요구된다.

Abstract The purpose of this study is to establish policy recommendations for the promotion of AI service impact assessment based on the definition of impact assessment and analysis of domestic and foreign AI service impact assessment cases. The direction of implementation was analyzed based on the case of impact evaluation promoted in various fields at home and abroad and the case of impact evaluation at home and abroad of artificial intelligence services. As a step-by-step implementation plan, in the first stage, quantitative indicators such as AI level survey-based economic effects are developed, and in the second stage, information culture such as safety and reliability and artificial intelligence ethics described in the Framework Act on Intelligence Information, social, economic, information protection, and people's daily lives are prepared. In the third stage, discussion on detailed metrics and methods will be expanded and impact assessment results will be evaluated. This study requires analysis through various participants such as policy designers, artificial intelligence service developers, and civic groups in the future.

Key Words : Artificial Intelligence, AI Service, National Strategy for AI, Impact Assessment

*정회원, 한국지능정보사회진흥원
접수일자 2022년 8월 8일, 수정완료 2022년 9월 8일
게재확정일자 2022년 10월 7일

Received: 8 August, 2022 / Revised: 8 September, 2022 /
Accepted: 7 October, 2022

*Corresponding Author: shinsy@nia.or.kr
National Information Society Agency, Korea

I. 서론

인공지능(Artificial Intelligence, AI)은 우리의 삶에서 점점 더 흔하게 등장하고 있다. AI 시스템은 다양한 범위의 복잡한 문제에 대한 통찰력을 얻고 정보에 기초한 결정을 내리는 데에 도움을 주고 있다.

AI의 혜택을 얻기 위해서 사회는 이러한 AI 시스템들의 리스크를 줄이고 완화해야 할 필요가 있다. 이에 AI의 리스크 평가와 영향평가가 강조되고 있으며 미국, 유럽 연합(EU) 등 정부 및 정부간 조직들은 AI 시스템의 사용으로 나타나는 리스크를 완화하고 사회가 AI의 혜택을 진정으로 즐길 수 있도록 보장하는 접근 방식을 고려하고 있다.

국내에서는 인공지능 국가전략('19)에서 인공지능에 의한 산업 및 사회의 역동적 변화에 대응하는 비전과 정책 방향을 공표하였다^[1].

산업발전과 경제성장을 향한 인공지능 전략을 추진하는 한편, 인공지능의 사회적 수용성과 긍정성을 극대화하고 발생 가능한 역기능을 최소화하기 위한 사회정책을 병행하는 전략이 담겨 있다. 또한, 지능정보화 기본법 제 56조에 인공지능 서비스를 포함한 지능정보서비스의 활용 확산으로 사회경제문화 및 국민의 일상생활에 미치는 영향에 대하여 평가하는 '사회적 영향평가'를 규정하고 있다. 그 추진에 대한 근거는 마련하였으나 인공지능 영향평가를 위한 구체적인 대응 방안 모색과 정책 수립이 이루어져야 한다. 이를 위해 인공지능 서비스 영향평가 활용을 위한 국내외 관련 연구의 다양한 관점 분석을 통해 인공지능 서비스 영향평가에 대한 현실성 있는 정책 모색이 필요하다.

II. 이론적 배경 및 현황

1. 인공지능

인공지능(AI)은 기계가 경험을 통해 학습하고 새로운 입력 내용에 따라 기존 지식을 조정하며 사람과 같은 방식으로 과제를 수행할 수 있도록 지원하는 기술이다. 인공지능(AI)은 학습, 문제해결, 패턴 인식 등과 같이 주로 인간 지능과 연결된 인지 문제를 해결하는 데 주력하는 컴퓨터 공학 분야에 속한다. 인공지능(AI)이라는 용어는 1965년에 처음 등장하였지만, 데이터의 양적 증가, 첨단 알고리즘, 컴퓨팅 파워와 스토리지가 개선된 오늘날 매우 활발히 연구되고 있는 분야이다. 2016년 구글의 알파

고를 이용한 바둑시합을 통해 인간 수준의 학습 추론 가능성을 제시하며 관심을 받기 시작하였다. 2000년대 들어 컴퓨팅 파워의 성장, 우수 알고리즘의 등장, 네트워크 발전에 따른 데이터 축적 등으로 급격히 진보하고 있다^[1].

2. 인공지능 서비스

인공지능 서비스의 개념을 정의해 보면, 인간의 지각, 이해, 판단, 학습 등에 대한 지적인 능력을 기기에 구현하고 빅데이터 등을 기반으로 분석, 예측하여 고객별 맞춤형 서비스를 제공하는 것을 의미한다. 인공지능 서비스는 기본적으로 수집과 축적된 데이터 기반으로 시작되고 각 데이터를 제공 서비스에 따른 알고리즘을 만들어 제공하는 원리를 가지고 있으며, 그 중심에서는 지속적인 데이터 피드백을 통한 딥러닝 프로세스를 보유하고 있다. 인공지능 기술은 의료, 금융, 제조, 교통, 게임, 예술, 범죄예방 등 다양한 산업에 접목되어 서비스되고 있으며 부가가치를 창출할 수 있다.

표 1. 인공지능 기술 분류

Table 1. Classification of artificial intelligence technologies

| 중분류 | 소분류 | 기술정의 |
|-------|----------|---|
| 학습 지능 | 머신러닝 | 지식, 기능 판단 등을 데이터분석, 시행착오, 기존지식 활용 등을 통해 학습하는 기술 |
| | 추론/지식 표현 | 기계가 이해할수 있는 형태의 지식 표현 및 기존 지식으로부터 새로운 사실을 추론하는 기술 |
| 단일 지능 | 언어지능 | 인간의 언어인지기능을 모사하여 텍스트 및 대화체 문장을 분석, 이해, 생성하는 기술 |
| | 시각지능 | 영상에서 사물의 위치와 내용(속성)을 이해하고 움직이는 행동(사건, 원인)을 이해하는 기술 |
| | 청각지능 | 인간의 청각기능을 모사하여 음향, 음성 등 소리를 인식, 분석, 이해, 표현하는 기술 |
| 복합 지능 | 행동/소셜 지능 | 공간을 인지하고 움직임을 제어하며 사회적 협업이 가능한 지능 |
| | 상황/감정 이해 | 주변 환경, 상황, 맥락, 인간의 감정 등을 다양한 센싱 정보로부터 인식, 분석, 이해하는 기술 |
| | 지능형 에이전트 | 특정한 목적을 위해 사용자를 대신해서 작업을 수행하는 인공지능 |
| | 범용 인공지능 | 인간이 할 수 있는 어떠한 지적인 업무도 수행할 수 있는 인공지능 |

국내 인공지능 주요 서비스는 제조, 헬스케어, 모빌리티, 공공·법률, 교육·고용, 금융, 안전, 문화, 광고 등의 인공지능 기술을 기반으로 한 서비스를 개발하고 있다. 서비스에 기반이되는 인공지능 기술은 학습지능, 단일지능, 복합지능으로 나눌 수 있으며 표1과 같이 각 분류별

세부 기술로 나누어 개발되고 있다.

인공지능 서비스는 기반 서비스로서 모든 분야에 적용될 수 있고 제공하고자 하는 분야 및 서비스의 목적에 따라 실시간 기계가 자동학습을 할 수 있는 환경을 제공하는 것이 중요하다.

국내 기업의 AI 도입은 여전히 2.5%^[2]에 불과하며, AI 제품 및 서비스를 잘 활용하고 있는 국민은 20%에 불과하다는 통계자료가 있다^[1,2]. 또한, AI 기술에 대한 국민의 신뢰도는 40.5%에 불과해 AI 서비스 체감효과는 아직 미흡한 상황으로 AI 서비스에 대한 혁신 방향 모색이 필요한 시점이다. 반면 글로벌 테크기업은 막대한 투자 비용을 감수하고 AI를 기반으로 기존 서비스 개선 및 신서비스 확장하고 있으며 AI 도입 적극성 수준에 따라 AI 기업의 이윤기여도에 격차가 심화될 것이다^[3].

3. 인공지능 위험 관리

Wharton School의 인공지능의 위험과 거버넌스 연구에 따르면, 펜실베이니아 대학 경영대학에서는 AIRS (Artificial Intelligence Risk and Security) 프로그램을 2019년부터 진행하고 있는데, 인공지능의 잠재적 위험은 인공지능 시스템 자체에서 발생하는 위험, 인공지능 시스템 사용에 따른 위험, 인공지능 시스템 거버넌스로 인한 위험으로 구분하고 있다.

EU는 인공지능을 위험도에 따라 금지, 고위험, 제한된 위험, 최소 위험으로 분류하여 의무를 부과하는 인공지능 규제(안)이 제안되었다^[4]. 인공지능 서비스는 개인에 대한 복잡한 데이터 흐름의 사용, 개인과 그들의 권리

에 대한 직접적인 영향 조직 및 생태계 전반에 걸친 조정 등 이는 이를 실현할 때 상당한 공통 위험 요소를 극복해야 한다.

인공지능 서비스 관련 등장 가능한 위험 요인들은 신뢰, 윤리, 거버넌스, 데이터, 지식 등이 존재한다.

UC Berkeley(2021)에서는 AI리스크를 중심으로 영향평가를 함께 어떻게 고려해야 하는지 분석하였다. 다섯 가지 문서 캐나다 정부의 자동 의사 결정 지침, 독일 데이터 윤리 위원회(Data Ethics Commission)가 제안한 2019 프레임워크, 2021년 4월 유럽 위원회가 제안한 법적 프레임워크, 뉴질랜드 정부가 자국 정부 기관 사용 목적으로 수립한 아오테아로아 뉴질랜드 알고리즘 헌장 (Algorithm charter for Aotearoa New Zealand), 샌프란시스코 도시 및 자치주 협력 조직인 고벡스(GovEx), 하버드 데이터스마트(DataSmart), 데이터 커뮤니티 DC(Data community DC)가 함께 개발한 윤리와 알고리즘 툴킷(Ethics and Algorithms Toolkit)에서 다양한 수준의 리스크를 어떻게 포함하고 있는지 비교 분석하였다^[5].

다섯 가지 문서 모두 AI 시스템의 아웃풋에 의한 영향의 특징을 고려하고 있으며, 대부분 인간의 기본권과 법적 권리, 인간의 물리적·정신적 복지에 미치는 영향을 검토하고 AI 시스템을 다양한 위험도 수준에 따라 분류하고 있다고 분석하였다.

4. 영향평가

가. 정책 영향평가

영향평가(impact assesment)는 공공 정책의 경제적, 사회적, 환경적 영향을 평가하는 증거 기반(evidence-based)의 절차로 사전적·사후적 의미의 평가를 모두 포함한다. 법률이나 정책의 개발 시 대상에게 미치는 영향을 분석하여 목표달성 가능성 및 예견되는 부정적 영향에 대한 제거방안을 확보하여 법률 및 정책의 성공적인 안착을 담보할 수 있어야 하며, 이를 위한 정책 분석을 사전예방적 평가, 또는 영향평가라고 한다.

노화준(1986)은 영향평가는 정책대안 또는 사업이 대상 집단이나 지역사회에 미치는 될 환경적 영향과 사회경제적 영향을 식별하고 추정함으로써 정책 결정과정에서 판단의 기초가 될 정책대안의 결과에 대해 유용한 정보를 제공하는 분석·평가"라고 하고 있다^[6].

유럽위원회(European Commission)에서 영향평가는 "정책선택의 이점과 약점을 실증적으로 제시함으로써

표 2. 인공지능 위험관리를 위한 거버넌스
 Table 2. Governance for AI Risk Management

| 구분 | 내용 |
|---------|--|
| 전담 조직 | • 조직 내 AI 거버넌스 프레임워크 구축은 이를 추진할 별도의 공식 조직을 구성하는 것으로 시작할 수 있으며, 이 공식 조직은 다양한 그룹과 부서를 대표 |
| 조직 활동 | • 모범사례 발굴 및 지식 공유, AI 시스템 사용을 위한 가이드 개발, 기존 워크플로우 지원 등의 활동을 위해 데이터 엔지니어와 LoB(Line of Business) 관계자 간 긴밀한 관계 구축 |
| 조직 역할 | • 공식 조직은 전문가 조직(CoE, Enter of Excellence), 위원회, 실무자 그룹 등 조직에 적합한 다양한 형태로 구성 가능 • AI 지원 이니셔티브 추진 시 모니터링 및 위험관리 체계나 내부 역할·책임 조정을 담당할 기구에 관한 논의 필요 |
| 모니터링 | • 중앙 모니터링 및 지속 보고 체계가 필수적이며, 여러 사람의 개입을 최소화하면 정확성, 일관성, 효율성 향상 가능 |
| 타사위험 관리 | • AI/ML 시스템에 서드 파티 애플리케이션 또는 데이터가 포함될 때, 타사 위험관리(TPRM)를 강화해 사전 테스트 필요 |
| 역할 | • AI 시스템 채택·구현과 관련하여 각 내부조직의 역할과 책임 논의할 조직 필요 |

정책 의사결정과정을 지원하며, 이를 통해 정책 행동이 수행되어야 할 수준과 범위를 확인하는 것"이라고 정의하고 있다.

OECD(2001)는 정책 영향평가를 정보에 기반을 두고 계획된 정책 수단의 비용, 결과, 부작용 등을 평가하는 활동으로 규정하고 정책효과 향상을 위해 필요한 도구임을 강조했다. 전통적인 국가 정책 과정에서 정당한 권한과 절차를 거칠 경우, 해당 제도 및 정책은 신뢰성이 확보되는 것으로 간주 되었고, 이에 대하여 사후적 평가가 주로 이루어졌다^[7]. 그러나 이러한 정당성과 신뢰성을 확보한 정책이 광범위한 부정적 영향 또는 회복 불가능한 손해를 발생시키게 되어 이를 사전에 예방하고, 긍정적 효과를 극대화하기 위하여 영향평가제도가 활성화되었다. 이는 영향평가의 정책 분석 수단으로서의 의의를 가진다.

정책은 그 속성상 '영향의 양면성'을 지니고 있으며, 이로 인하여 정책으로 인하여 이익을 받는 국민(집단)과 함께, 반대로 불이익을 받는 국민(집단)이 필연적으로 발생하므로 부정적인 영향을 받는 대상이 많거나 그 정도가 심한 경우 국가 및 사회 전체에 갈등을 유발할 수 있다. 그러므로 정책 등이 결정되기 이전에, 검토 중인 정책대안이 초래할 결과를 예측하고 예상된 결과가 국가와 사회에 미칠 영향을 제시하여 정책결정자의 결단이 보다 합리적이고 바람직할 수 있도록 지원하는 사전영향평가(ex ante impact evaluation)가 유용성을 가질 수 있다.

정책 분석은 과거지향적(사후적) 정책 분석과 미래지향적(사전적) 정책 분석으로 구분할 수 있다. 이중 사전적 정책 분석은 법률의 제·개정 및 정책 행위가 시작되기 이전 분석을 통한 지식의 생산과 전환이 이루어지는 것이다. 영향평가는 특정 사업이나 정책의 사회적·경제적·환경적 영향을 사전에 검토하여 미래에 가져올 결과들에 대한 바람직한 대응 방안을 제시함으로써 보다 합리적인 정책 결정을 유도하는 유용한 정책 분석·평가 기법으로서 의의가 있다^[6].

나. 환경 영향평가

우리나라에서 가장 대표적인 영향평가의 영역은 환경영향평가(Environmental Impact Assessment)로서 환경정책기본법을 통해 환경영향평가를 규정한 후, 1993년 6월 환경영향평가법을 제정하여 최초의 독립된 영향평가 법제를 도입하였으며, 환경영향평가법 제정 이후 다수의 영향평가가 도입되어 운영되고 있으며 이외에도

다수의 영향평가제도가 도입 논의를 계속하고 있다. 환경영향평가제도에 대한 개선책으로 '사회영향평가'(Social Impact Assessment) 제도의 도입이 꾸준히 제안되어 왔다^{[6][7]}. 사회적 영향평가의 개념은 다양하게 정의된다. 제안된 정책·계획·프로그램·사업 등에 의하여 사람들의 생활양식, 노동형태, 여가방식, 조직구조, 관계 변화 등 사회에 일어나는 결과들 - 이것을 '사회영향'이라고 한다 - 을 예측·평가하는 제도라고 정의하기도 하고, 제안된 정책·계획·프로그램·사업 등이 "가시적인 영향을 드러내기 전인 계획단계에서 사회 영향을 예측·평가하는 것", 또는 "개인, 집단, 지역사회에 미칠 수 있는 잠재적 영향을 평가하는 절차"라고 기술되기도 한다^[7].

환경영향평가는 사전환경성 검토' 및 '환경영향평가' 제도가 '환경정책기본법'과 '환경영향평가법'으로 각각 운용되고 있어 처리 절차가 복잡하고 적용에 일부 혼선도 있는 등의 문제 발생했다. 2012년 7월부터 개정법에 따라 '전략환경영향평가', '환경영향평가', '소규모 환경영향평가'로 나누어 진행하였다^[6].

다. 개인정보 영향평가

개인정보 영향평가는 개인정보처리 시스템에 위주의 평가이며 인공지능 영향평가는 데이터, 알고리즘, 서비스 개발 및 배포, 사용자, 정책당국 등 고려 요소가 많으며, 생태계적 접근방식이 필요하다^{[5],[6]}.

표 3. 개인정보 영향평가 (PIA)

Table 3. Personal Information Impact Assessment

| 구분 | 내용 |
|----------|--|
| 개념 | <ul style="list-style-type: none"> 개인정보파일을 운용하는 새로운 정보시스템의 도입이나 기존에 운영 중인 개인정보 처리시스템의 중대한 변경 시 시스템의 구축·운영·변경 등이 개인정보에 미치는 영향(impact)을 사전에 조사·예측·검토하여 개선방안을 도출하는 체계적인 절차 |
| 목적 및 필요성 | <ul style="list-style-type: none"> 개인정보 처리가 수반되는 사업 추진시 해당 사업이 개인정보에 미치는 영향을 사전에 분석하고 이에 대한 개선방안을 수립하여 개인정보 침해사고를 사전에 예방 |
| 평가 대상 | <ul style="list-style-type: none"> 공공기관(개보법시행령 35조) (5만명 조건) 5만명 이상의 정보주체의 민감정보 또는 고유식별정보의 처리가 수반되는 개인정보파일 (50만명 조건) 해당 공공기관의 내부 또는 외부의 다른 개인정보파일과 연계하려는 경우로서, 연계 결과 정보주체의 수가 50만 명 이상인 개인정보파일 (100만명 조건) 100만 명 이상의 정보주체 수를 포함하고 있는 개인정보파일 (변경 시) 영제35조에 근거하여 영향평가를 실시한 기관이 개인정보 검색체계 등 개인정보파일의 운용체계를 변경하려는 경우, 변경된 부분에 대해서는 영향평가 실시 |



그림 1. 개인정보 영향평가 프로세스
 Fig. 1. Personal Information Impact Assessment Process

라. 기술 영향평가

기술영향평가는 과학기술이 가지는 역기능을 최소화하고 순기능을 극대화함으로써, 과학기술에 대한 사회적 인식과 기술 수용성을 높이고자 목적을 가지는 것으로 이해해볼 수 있다^[8].

기술영향평가는 결국 기술적 발달이 미치는 영향을 사전에 예측하고 그에 관해 선제적으로 대응함으로써 안정적인 과학기술의 발전을 도모하는 정책적 평가도구라고 할 수 있다. 과학기술평가는 1990년대부터 시작되었으며 현재는 1년 단위로 수행하며 과학기술기본법에도 명시되어 있다.

마. 기타 영향평가

영향평가의 효과성 측면에서 최근 도입되고 있는 다수의 영향평가는 규범적 평가로서의 성격을 가지고 있으며, 사전 평가를 통하여 해당 분야에 미치는 부정적 영향을 최소화하고 정책이 효과적으로 집행될 수 있도록 지원하는 역할을 하고 있다. 규범적 영향평가의 경우 평가

표 4. 영향평가의 종류
 Table 4. Types of Impact Assessment

| 구분 | 영향평가 종류 | 효과 | 장단점 |
|-----|------------------------------------|---|---|
| 규제형 | 환경영향평가, 교통영향평가, 재해영향평가, 건강영향평가 등 | 평가의 결과실시 및 시행계획 등의 허가인가승인 면허 등 행정행위에 영향을 미침 | 제재를 통하여 효과적 평가 목적 달성 가능. 평가에 따른 사회적 비용이 과다하며 규제심화로 인한 진입장벽 발생 |
| 규범형 | 성별영향평가, 고용영향평가, 개인정보영향평가, 기술영향평가 등 | 체계적인 평가를 통해 정책 등에 있어 합리적인 개선 유도 | 제도의 의의 및 목적, 가치확대 용이. 평가의 실효성 문제 발생 |

대상 사업이 명확하게 정해진 것이 아니라 각 법률안 및 정책의 주관부처 또는 지방자치단체가 관련 정책에 대한 특정 영향평가를 요청하는 경우 이루어지며, 지방자치단체의 개별적 영향평가는 지자체마다 다른 대상 범위를 가지고 있다^[6].

국내에서 이루어지는 주요 영향평가는 아래 표5와 같이 정리될 수 있다.

표 5. 국내 주요 영향평가
 Table 5. Major Impact Assessment in Korea

| 구분 | 부처 | 목적 |
|----------|-----------|---|
| 환경영향평가 | 환경부 | 자연, 생활, 사회, 경제, 환경영향 예측 및 대책 강구(인구영향평가, 교통영향평가, 재해영향평가의 상호 중복문제 등으로 환경영향평가로 변경-대체-폐기) |
| 건강영향평가 | 환경부 | 주요 정책 또는 개발사업이 주변 지역 주민의 건강에 미칠 수 있는 영향을 예측, 평가 |
| 성별영향평가 | 여성가족부 | 실질적인 성평등 정책 실현을 위해 법령과 정책에 성적 불평등이 있는지 여부를 자체 평가하여 여성부에 제출하면 여성부가 종합 평가하는 제도 |
| 고용영향평가 | 고용노동부 | 국가 지자체의 주요사업 정책 및 법 제도가 고용에 미치는 영향을 분석 평가하여 고용 친화적 정책추진을 지원 |
| 규제영향분석 | 국무총리실 | 정부 각 부처가 규제를 신설하거나 강화하는 내용의 정책을 입안할 경우 입법예고 단계에서 규제영향분석서를 작성하여 이해관계자의 의견을 수렴하는 제도 |
| 경쟁영향평가 | 공정거래위원회 | 경쟁 영향평가는, 정부나 여러 규제기관이 특정한 목적을 달성하기 위하여 도입하고자 하는 새로운 규제 또는 이미 도입한 규제가 경쟁에 어떠한 영향을 미치는지를 분석평가 |
| 기술영향평가 | 과학기술정보통신부 | 국가 R&D 사업 기획시 미래의 신기술이 가져올 각종 영향과 부작용을 KISTEP 주관으로 사전에 평가하는 제도 |
| 문화영향평가 | 문화체육관광부 | 국가 또는 지방자치단체의 각종 정책-제도-사업에 대해 경제적 파급효과 외에 문화적 타당성을 검증하도록 하는 제도 |
| 인권영향평가 | 국가인권위원회 | 공공정책 등이 국민생활에 중대하고 광범위한 갈등을 유발할 우려가 있다고 판단되는 경우 공공기관의 장이 갈등 영향분석을 실시하거나, 동사업을 시행하는 민간사업자에게 갈등 영향분석을 실시하도록 하는 제도 |
| 아동정책영향평가 | 보건복지부 | 국가와 지방자치단체가 아동 관련 정책이 아동복지에 미치는 영향을 분석하고 평가하여 그 결과를 정책의 수립 및 시행에 반영 |

III. 인공지능 영향평가

1. 인공지능 영향평가의 정의

인공지능 영향평가를 협의 측면에서 논의하면, AI가 윤리적이고 합법적으로 사용되도록 기회와 위험을 초기

단계에서 등장하는 부정적인 문제 발생을 예방 목적이 다. 즉 인공지능 서비스는 (1) 책임성, (2) 문화에 적합한 윤리성, (3) 합법성을 보유하고 있어야 하므로 이를 보유하지 않을시 등장하는 위험을 파악하여 대응하고 완화하려는 목적으로 진행한다.

광의의 측면에서 보면 (1) 한국의 지능 정보화 기본법에서의 AI 영향평가, (2) AI 서비스 안전성 및 신뢰성, 정보문화, 사회·경제, 정보보호, 일상생활, (3) 사전, 사후 평가한 혁신적 기술의 영향 프레임워크 등을 논의 할 수 있다^{[1], [7], [8]}.

지능정보화 기본법 제 56조는 영향평가 주체, 대상, 주요 평가항목, 일부 절차 등을 규정하고 있다. 그러나 일반적인 평가 절차인 평가 대상·범위 결정, 조사·평가 실시, 결과 분석 및 공개, 의견 수렴 등이 규정되어 있지 않다. 포괄적, 평가항목을 열거하여 인공지능 서비스의 안정성, 신뢰성, 기타 영향평가 결과에 따라 조치 및 권고가 가능한 환경을 제공하고 있다.

2. 인공지능 영향평가 방법

인공지능 기술의 급속한 인공지능 기술의 급속한 발전으로 다양한 분야에서 인공지능 서비스가 활용확산됨에 따라 인공지능 서비스의 국민생활 전반에 미치는 영향력 증가하고 있다. 이에 이에 인공지능 서비스의 사회·경제·문화 및 국민의 일상생활 등에 미치는 영향에 대하여 평가 및 종합적 분석 필요하다. 또한, 여러 연구자들이 우려하고 있는 인공지능 서비스의 부정적 영향을 최소화하거나 방지하기 위한 제언을 통해 인공지능 서비스의 안전성 및 신뢰성 향상 등에 기여함을 고려하여 인공지능 영향평가를 설계하여야 한다. 현재 지능정보사회기본법 56조의 항목을 기준으로 한 영향평가 세부 분류는 크게 5개 카테고리로 정리할 수 있다. 첫째, 인공지능 서비스의 안전성 및 신뢰성 부분, 둘째, 정보격차 해소, 사생활 보호, 인공지능 윤리 등 정보문화에 미치는 영향, 셋째, 고용·노동, 공정거래, 산업 구조, 이용자 권익 등 사회·경제에 미치는 영향, 넷째, 정보보호에 미치는 영향, 다섯째, 국민의 일상생활에 미치는 영향으로 법에 정의하고 있으며 각 파트별로 세부 측정지표나 방식 등의 논의를 확대해 나갈 수 있다^{[9], [10], [11]}.

또한, 대상이 되는 서비스를 주기적으로 선정하여 그 영향을 평가하여야 하는데 새로운 서비스일 뿐만 아니라, 향후 미래에 국민생활에 높은 파급효과가 있을 것으로 예상되어 영향평가의 유용성이 높을 것으로 예견되는 서비스, 사회적 관심도가 높은 서비스 등의 기준을 가지

고 다양한 이해관계자의 의견 수렴을 통해 서비스가 주기적으로 선정될 필요가 있다.

일반적인 영향평가는 사전평가와 사후 평가 방법으로 분류될 수 있으며, 인공지능 서비스 도입 이전에 예상되는 영향성 평가를 통해 통제관리를 강화하는 측면으로 사전 평가를 수행하고, 인공지능 서비스를 도입 이후에 등장하는 위험요소를 포함한 다양한 영향성을 평가하는 경우로 사후 평가를 진행할 수 있다^[6, 7, 11].

표 6. 사전 평가와 사후 평가

Table 6. Pre impact Assessment and Post impact Assessment

| 구분 | 내용 | 수행방안 |
|---------|---|--------------------------------------|
| 사전 영향평가 | • 인공지능 서비스 도입 이전에 예상되는 영향성 평가를 통해 통제관리를 강화하는 측면 | • 인증, 허가제 도입, 제도적 규정, 가이드라인 도입 등이 있음 |
| 사후 영향평가 | • 인공지능 서비스를 도입 이후에 등장하는 다양한 영향성을 평가하는 경우 | • 소비자 보호 활동, 책임 소재 규명 |

3. 인공지능 수준 조사

국내외에서는 인공지능 정책 현황 분석, 다양한 분야의 AI 성과를 측정하고 AI 경쟁력을 진단하기 위한 수준 조사를 진행하고 있다. 이는 AI 분야 발전을 위해 AI가 영향을 미치는 다양한 분야에 대한 데이터 기반의 객관적인 수준 진단과 이에 맞는 지표체계 필요하다. 기술이 직접적으로 영향을 미치는 산업·경제 분야뿐만 아니라 연구·교육, 정책·문화 등 사회 전반을 포괄하도록 지표를 설계하고 있다.

데이터 기반으로 진단한 AI 분야 수준을 바탕으로 관련 분야 지원 정책 수립의 근거를 마련하고, 정부와 기업의 의사결정을 지원할 수 있어서 스탠포드 등 여러 연구기관에서 조사 결과를 발표하고 있다. 스탠포드 AI 인덱스 주요 특징으로는 'AI 인덱스'를 기반으로 AI 관련 데이터를 한곳에 모으고 공유하는 협업 체계를 확립하고, 블룸버그(정책), 링크드인(일자리), 맥킨지(산업) 등의 전문가들이 지표 개발과 데이터 수집·분석에 참여한다는 점이다. 따라서 AI 전문가 협업 체계를 바탕으로 안정적이고 정확한 데이터 수집이 가능하며, 각 분야 전문가들의 심도 있는 분석 및 의견 수렴이 가능하다는 장점이 있다. 수준 진단 결과에 따라 집중 투자 분야 등을 타 국가와도 비교할 수 있다^[12].

표 7. 스탠포드 AI 인덱스
 Table 7. Stanford AI Index

| 구분 | 내용 |
|----|--|
| 목적 | • 1년의 AI 분야 변화를 데이터 기반으로 측정하여, AI 관련 정책적 의사결정 지원 |
| 기관 | • 스탠포드 HAI 연구소(미국) (보고서) 5개 분야, 120여 개의 지표 (글로벌 AI 활동성 도구) 연구개발, 경제분야의 23개 지표 |
| 특징 | • 데이터와 출처, 수립 절차를 모두 공개하는 오픈 프로젝트 • 국가 간 비교가 가능한 지표를 선별하여, 국가 순위 도출이 가능한 계산 도구 제공 |

4. 인공지능 영향평가 사례

가. NIST의 AI RMF

NIST에서는 인공지능 위험관리 프레임워크 개발을 2021년부터 진행중에 있다^[13].

표 8. NIST의 인공지능 위험의 분류 및 위험관리 프레임워크
 Table 8. NIST's AI Risk Classification and Risk Management Framework

| 구분 | 내용 |
|---------------------|--|
| 기술적 속성의 시스템적 위험 | • 시스템 설계자/개발자들이 사용하는 일반적으로 사용하는 평가기준으로서 시스템의 정확성, 신뢰성, 견고성, 보안성의 위험 |
| 사회 기술적 속성에서의 인지적 위험 | • AI 시스템이 어떻게 인지되는가에 따른 위험, 즉 AI 모델에 대해 인간이 지각, 예를 들면 AI 시스템의 투명성, 설명가능성, 해석가능성, 프라이버시, 안전 등 AI 시스템에 대한 인간의 주관적 평가, 판단, AI 결과에 대한 인간의 해석이 필요한 상황에서의 위험 |
| 신뢰성 정책과 사회적 위험 | • 책임성, 공정성, 정의, 공정 등 사회적 가치의 손상으로부터 발생하는 상황 의존적 위험 |

AI 시스템에 신뢰성(trustworthiness)을 제공하기 위해 2022년 3월부터 AI RMF를 개발하기 위한 컨셉페이퍼를 발표하였다. AI RMF는 2023년 초 정식 문건 발간을 목표로 현재 오픈 방식으로 진행하고 있다. 핵심기능, 프로파일, 구현 가이드로 구성되어 있다.

AI 관련 제품, 서비스 및 시스템을 설계, 개발 사용, 평가하면서 발생할 수 있는 위험을 해결하는데 참조할 수 있는 자료들의 모음이다.

나. EU의 SHERPA

EU의 SHERPA 프로젝트는 다음과 같은 목적으로 진행되고 있는 프로젝트이다. 스마트 정보 시스템(SIS, 인공지능과 빅데이터 분석의 조합)이 윤리 및 인권 문제에 영향을 미치는 방식에 대한 이해를 조사, 분석 및 종합하

고 있다. 이 프로젝트는 SIS 문제를 이해하고 해결하는 새로운 방법을 개발하고 이해관계자와 함께 평가하는 권장사항 중 하나로 AI 영향평가를 위한 기본 모델 개발이 진행되었다. 그 구성요소로 AI 영향평가 수행 및 평가 프로세스, AI 영향에 대한 측정 및 매트릭, 기술 또는 응용 프로그램에 대한 위험 수준 결정, AI로 발생하는 수많은 윤리적 인권 문제에 대한 평가 지침을 제공하여야 함을 권고하고 있다^[14,15].

표 9. EU의 SHERPA 수행내용
 Table 9. EU SHERPA

| 구분 | 내용 |
|----|---|
| 1 | 개념의 명확성 AI 및 디지털 기술에 대한 적절하고 명확한 정의 사용 |
| 2 | AI 영향평가 AI 영향평가를 위한 기본 모델 개발 |
| 3 | 디자인에 대한 윤리 EC 자금 지원 프로젝트 서비스 디자인 윤리 |
| 4 | AI 및 윤리교육 AI의 윤리 및 인권을 포함하는 훈련 및 교육 |
| 5 | 표준화 AI 윤리에 대한 연구 결과를 표준화 |
| 6 | 규제 프레임워크 AI에 대한 규제 프레임워크 및 시행 매커니즘 개발 |
| 7 | AI에 대한 유럽기관 AI를 위한 독립적인 유럽 연합 기관을 설립 |
| 8 | AI 윤리 담당자 조직 내 AI(윤리) 담당자 역할 설정 |

다. 영국 알고리즘 영향평가

영국에서는 헬스케어 분야에 대해서 AI 알고리즘에 대한 영향평가 모델을 발표하였다. 알고리즘 영향평가(AIA)는 이러한 위험을 완화하기 위해 AI 시스템이 줄 수 있는 사회적 영향을 사전에 평가하는 데 사용되는 도구(AI 시스템에 대한 사후 지속적인 모니터링에도 사용 가능)로서 영국 국민건강서비스(NHS)는 최근 AI와 데이터 전문 연구 기관인 Ada Lovelace Institute에서 개발한 AIA 모델을 사용하여 세계 최초로 헬스케어 분야에서 AIA 시범적용을 한다고 발표하였으며 NHS는 NHS AI Lab의 국가의료영상플랫폼(NMIP) 데이터에 액세스하여 AI 시스템을 개발하려는 회사는 누구든지 AIA를 수행해야 한다는 것을 명확히 하고 있다^[16,17].

라. 윤리영향평가

2021년 11월에는 유네스코에서 인공지능 윤리권고사항이 발표되었으며 이는 4대 가치 및 10대 원칙을 발표하였다. 여기에는 AI 시스템의 위험을 식별 평가하고 모니터링 조치를 개발하기 위한 윤리 영향평가 프레임워크 도입의 필요성이 포함되었다. 또한, AI 기반 결정이 인간

의 의사결정에 미치는 사회적 심리적 영향을 조사하여야 한다는 권고도 포함되었다. 또한, 2022년 5월 인권위원회에서 발표한 인공지능 관련한 인권 영향평가에서도 인공지능 서비스 추진시 인권침해 요소에 대해서 고려하여야 한다는 권고를 하였다^[9,10,18].

마. ECP의 AIIA

네덜란드ECP의 AIIA(Artificial Intelligence Impact Assessment)는 다음과 같은 내용으로 기업이 인공지능 영향성 평가를 위해 7단계로 수행하는 방안을 제시하였다. 캐나다 정부에서는 행정주체(정부) 스스로 행정 결정

표 10. ECP의 AIIA의 서비스 영향 수행 프로세스
Table 10. ECP's AIIA's Service Impact Implementation Process

| 구분 | 내용 |
|-----------------|---|
| AIIA 수행의 필요성 결정 | 1. AI가 새로운(소설) 영역에서 사용되니까? 2. 새로운 형태의 AI 기술이 사용되니까? 3. AI는 자율성이 높은가? 4. AI가 복잡한 환경에서 사용되니까? 5. 민감한 개인 데이터가 사용되니까? 6. AI가 사람이나 단체에 심각한 영향을 미치거나 법적 절차를 초래하는 결정을 내립니까? 7. AI가 복잡한 결정을 내립니까? |
| 1단계 | AI 애플리케이션 설명 1. 지원서 및 지원서의 목적 설명 2. 목표를 달성하기 위해 어떤 AI 기술이 사용되는지 설명 3. 애플리케이션 컨텍스트에서 사용되는 데이터 설명 4. 애플리케이션에서 어떤 액터가 역할을 하는지 설명 |
| 2단계 | AI 애플리케이션의 이점 설명 1. 조직에 어떤 이점이 있습니까? 2. 개인에게 어떤 혜택이 있습니까? 3. 사회 전체에 어떤 이점이 있습니까? |
| 3단계 | 애플리케이션이 신뢰할 수 있고 안전하며 투명합니까? 1. AI 행동의 신뢰성을 보장하기 위해 어떤 조치가 취해졌습니까? 2. AI의 안전성을 보장하기 위해 어떤 조치가 취해졌습니까? 3. AI 행동의 투명성을 보장하기 위해 어떤 조치가 취해졌습니까? |
| 4단계 | 목표와 목표에 도달하는 방법이 윤리적이고 법적으로 정당합니까? 1. 내 AI 애플리케이션에 관여하거나 영향을 받는 행위자는 무엇입니까? 2. 이러한 가치와 이해관계가 법령에 규정되어 있습니까? 3. AI 배포의 맥락에서 어떤 가치와 관심이 중요한 역할을 합니까? |
| 5단계 | 애플리케이션이 신뢰할 수 있고 안전하며 투명합니까? 1. AI 행동의 신뢰성을 보장하기 위해 어떤 조치가 취해졌습니까? 2. AI의 안전성을 보장하기 위해 어떤 조치가 취해졌습니까? 3. AI 행동의 투명성을 보장하기 위해 어떤 조치가 취해졌습니까? |
| 6단계 | 고려 사항 및 평가 |
| 7단계 | 문서화 및 책임 |
| 8단계 | 주기적으로 검토 |

에 사용되는 자동화 결정 시스템에 대하여 알고리즘 영향평가(AIIA)를 거칠 것을 필수 요건(의무사항)으로 규정하고 있다. 자동화 결정 시스템이 미치는 영향의 정도는 4단계로 구분하여 평가되고, 1단계부터 4단계의 구분은 AI 시스템이 내린 결정(판단)이 개인이나 공동체의 권리, 건강, 복지, 경제적 이익, 생태계의 지속 가능성 여부에 대하여 어느 정도의 영향을 미치는지에 따라 구분한다^[19, 20].

바. IEEE의 WIA

IEEE에서는 웰빙 서비스에 지능형 시스템의 영향성 평가 사례를 통해 수행 프레임워크 파악하기 위해 IEEE Well-Being Impact Assessment (WIA)를 진행 하였다. 영향평가를 위한 검토사항으로 활동 내용을 제시하고 그 활동 내용에 대한 지침과 체크리스트를 작성한다. 제시한 활동 지침을 충족할 수 있는 측정지표를 선정하고 측정지표에 맞는 평가 문항을 개발하려는 전략을 가지고 있다. WIA(Well-Being Impact Assessment) 및 웰빙 지표 대시보드 활동 및 단계에 대한 지침 내용을 다음 5가지로 정리하였다. 활동 1은 내부, 사용자 및 이해관계자 영향평가, 활동 2는 웰빙 지표 대시보드, 활동 3

| 구분 | 검토사항/체크리스트 |
|----------|--|
| 내부분석 | 1. AI는 무엇입니까? |
| | 2. 그것이 충족하거나 해결하는 문제는 무엇입니까? |
| | 3. 의도된 사용자 의도하지 않은 사용자 및 이해 관계자는 누구입니까? |
| 사용자 참여 | 4. 인간 행동에 대한 가능한 영향, 발생 확률은 무엇이며 인간 행동에 대한 부정적인 영향은 어떻게 고려되고 완화될까? |
| | 1. 사용자 행동에 어떤 이점이 있습니까? |
| 이해관계자 참여 | 2. 사용자 행동에 해가 되는 것은 무엇입니까? |
| | 3. 인간 행동에 대한 가능한 영향, 발생 확률은 무엇이며 인간 행동에 대한 부정적인 영향은 어떻게 고려되고 완화될까? |
| | 4. 의도하지 않은 사용자 및 영향 발생 확률은 무엇이며 인간 행동에 대한 부정적인 영향은 어떻게 고려되고 완화될까? |
| | 1. 이해 관계자의 특성에 대한 이점은 무엇입니까? |
| | 2. 이해 관계자의 특성에 대한 해는 무엇입니까? |
| | 3. 인간 행동에 대한 가능한 영향, 발생 확률은 무엇이며 인간 행동에 대한 부정적인 영향은 어떻게 고려되고 완화될까? |
| | 4. 의도하지 않은 사용자 및 영향 발생 확률은 무엇이며 인간 행동에 대한 부정적인 영향은 어떻게 고려되고 완화될까? |

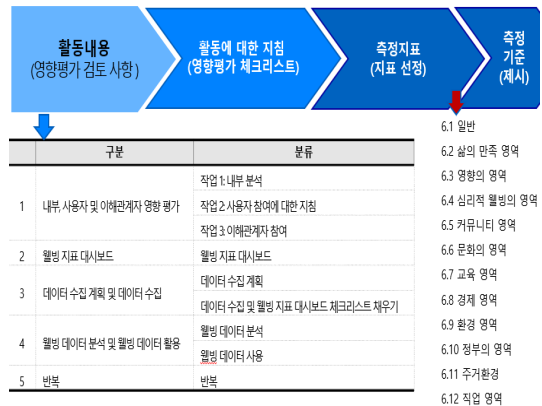


그림 2. IEEE 웰빙영향평가
Fig. 2. IEEE Well-Being Impact Assessment (WIA)

은 데이터 수집 계획 및 데이터 수집, 활동 4는 웰빙 데이터 분석 및 웰빙 데이터 활용, 활동5는 수행내용을 반복하여 진행하는 것이다^[21].

IV. 인공지능 영향평가 정책 제언

1. 인공지능 영향평가를 위한 접근 방법

본 연구에서 파악한 내용을 기반으로 다음과 같이 단계별 인공지능 영향평가 추진을 위한 접근방식은 서비스 분야별 측정항목 외에도 통합적 평가를 위해서는 정성, 정량 평가를 모두 실시하는 혼합적 방법론 바탕의 측정 지표 발굴이 필요하다. 또한, 타 영향평가 사례를 바탕으로 볼 때 영향평가 주체·대상·절차·방법 등을 구체화함으로써 영향평가의 합목적성 및 실효성을 확보할 수 있는 제도적 기반 마련이 중요하고 영향평가 대상 분야·서비스 선정, 평가지표 수립·적용, 평가 결과 공개 및 반영에 이르는 영향평가 체계 전반의 합리성·전문성·실효성 확보를 위한 방법론 모색도 필요하다. 아래 표11은 사례를 바탕으로 한 영향평가에 대한 고려 요소를 요약하였다^[10, 11, 22].

표 11. 인공지능 영향평가를 위한 고려 요소
 Table 11. Factors to Consider for AI Impact Assessment

| 구분 | 내용 |
|--|--|
| 생태계 관점에서의 접근 필요 | <ul style="list-style-type: none"> 인공지능 서비스 개발자, 제공자, 사용자, 데이터 제공자, 정부 및 단체 등 다양한 이해관계자의 요구 사항 수렴 |
| 전통적인 정보보호가 아닌 융합보안 차원에서 레질리언스 관점에서 접근 필요 | <ul style="list-style-type: none"> 인공지능은 범용 기술이며 이는 금융, 의료, 자동차, ICT 시스템, 제조(산업제어 시스템) 등 모든 영역에서 사용됨으로 서비스 특성이 반영된 융합보안 차원에서 안전성을 점검 Security(정보 + IT 인프라) + Safety(인적자원(사용자, 일반 대중)의 안전 및 생명 + Physical(제품, 시설물 등) 전통적 보안이 예방 쪽에 초점을 맞추었다면 보안사고의 발생을 기정사실화하고 이에 대한 탐지와 대응 능력을 강조하는 레질리언스 역량 필요 |
| 인공지능 서비스 생애주기 전 단계에 걸친 위험관리 접근 필요 | <ul style="list-style-type: none"> 서비스 설계, 개발, 구현, 배포, 폐기 등 모든 단계에서의 안전성 점검 필요 서비스 설계 위험뿐만 아니라 서비스 사용에 따른 위험, 준거성 위험까지 고려 |
| 개인화된 인공지능 서비스로 인해 프라이버시 침해 가능성에 대한 접근 필요 | <ul style="list-style-type: none"> 전통적인 개인정보보호(정보에 초점)가 아닌 기본권으로서 사생활 보호 측면에서 확장된 영향평가 고려 필요. 신뢰성 정책과 원칙 (책임성, 공정성, 정의, 공정) 등 사회적 가치의 손상으로부터 발생하는 상향 의존적 위험에 대한 영향을 평가 필요 |

2. 정책 제언

추진을 위한 정책 제언으로 첫째, 어떤 측면을 중심으로 인공지능 서비스 영향평가를 진행할지에 대한 결정이 필요하다. 기술성, 시간적, 서비스 분야별, 우선순위 등의 결정 및 프로세스 정립이 시급하다^[17].

둘째, 영향평가의 긍정적 영향 중심의 정량평가는 국가 통계나 사례 분석 등을 통해 영향평가를 위한 지표 및 모델을 확립하여 AI 잠재력 최대화를 위한 구체적인 정책 제언이나 정책과제를 제시할 수 있는 기반을 마련할 수 있다. 사회적 이슈 중심의 정성평가는 선정한 서비스 분야의 AI 활용에 따른 위험관리를 위한 구체적인 법제 도적 과제를 도출할 수 있다. 정량평가와 정성평가 두 가지 방법을 혼합한 방법을 고려하는 것도 필요하다.

셋째, 일종의 '사후적 사회 영향평가'로서 위상을 갖는 현행법상 인공지능 서비스 영향평가는 사전적 위험 예방 차원에서 시행하는 각종 영향평가제도와 차이가 있으며, 영향평가 요소를 구체화하는 과정에서 본 제도 특유의 목적과 성격을 중요하게 고려할 필요가 있다. 여러 영향평가제도가 그러하듯이 인공지능 영향평가의 경우에도 실제로 평가 대상과 범위를 결정하고 영향을 조사·평가하는 전문조직의 설치가 필요하다. 또한, 평가의 정밀성과 실효성을 높이기 위해 민간 사업자 등 서비스 제공자가 참여하는 민관 협력의 거버넌스가 우선시되어야 하며, 무엇보다 광범위한 시민 의견 수렴을 위한 포럼 운영을 필수적으로 고려해야 할 것이다^[6].

넷째, 서비스 영향평가의 운영시 인공지능 서비스에 대한 특성이 고려되어야 한다는 점이다. 앞서 살펴본 인공지능 서비스 분류를 리스크를 기반으로 분류한 사례, 데이터의 특성으로 분류한 사례가 있었다^[5, 13]. 또한, 인공지능 서비스의 특성상 서비스 이후 계속해서 학습하고 알고리즘이 발전할 경우에 대해서는 주기적, 지속적 평가를 고려할 필요가 있다. 또한, 매년마다 특정 인공지능 서비스에 대한 영향평가를 진행할 경우 어떤 기준으로 서비스를 도출할지에 대한 우선순위 도출이 필요하여 서비스 생애주기와 같은 개발 생태계 또한, 고려되어야 할 것이다^{[22],[23]}.

3. 단계별 인공지능 영향평가

본 연구에서 파악한 내용을 기반으로 다음과 같이 단계별 인공지능 영향평가 방안을 제시하였다.

표 12. 단계별 영향평가 수행에 따른 해결 과제
Table 12. Challenges to step-by-step impact assessment

| 단계 | 구분 | 내용 | 해결과제 |
|-----|-----------------------------------|---|--|
| 1단계 | AI 인덱스 확장 기반 영향평가 | <ul style="list-style-type: none"> AI 수준평가 + 사회, 경제 영향(고용, 노동, 경제적 지표 등) 국가 차원에서 AI 수준/역량 측정과 자료 확보 목적 | <ul style="list-style-type: none"> 측정지표의 개선 사항과 추가 지표 발굴 지표 측정 방법 정의 |
| 2단계 | AI 서비스 안정성을 위한 영향평가 | <ul style="list-style-type: none"> AI서비스 안정성 평가 방향 포함 인공지능 서비스 제공기관 등에 대한 영향 평가 실시 특정 서비스에 대한 사전적 영향평가 | <ul style="list-style-type: none"> 국내 상황에 맞는 지표 발굴 제도적 방안 마련 |
| 3단계 | AI 서비스 신뢰성을 위한 영향평가: 윤리, 이용자 권익 등 | <ul style="list-style-type: none"> 윤리 영향평가(Ethics by Design), 알고리즘 영향평가 사전적 영향평가와 사후적 영향평가 체계 사전영향평가 수행 결과를 지표로 반영하여 정책 차원에서 영향평가 보완 | <ul style="list-style-type: none"> 윤리적 딜레마에 적용할 수 있는 방안과 지표 마련 지표 평가를 위한 제도적 방안 마련 |

1단계는 AI 인덱스 확장 기반 영향평가로서 기존의 연구 된 내용을 기반으로 국내에 적합한 방안 연구를 통해 구체적으로 시장에 적용하여 평가할 수 있는 대안을 마련하는 것이다. 또한, AI 영향평가는 정책 또는 사업의 허가·인가·승인·면허 등 행정행위를 위한 것이 아닌, 정책의 합리적 개선을 유도하기 위한다는 취지를 고려할 때 ‘규범형’이 바람직하다.

2단계는 AI 서비스 안정성을 위한 영향평가로서 인공지능 서비스의 설계/개발 단계에서 사이버 레질리언스 역량과 프라이버시 보호 기능을 평가하는 것으로, 적용 대상은 공공기관에서 수행하는 인공지능 서비스를 대상으로 하되 민간기업의 경우, 인센티브를 제공하는 방식으로 간접적인 유도방안을 제시한다. 3단계는 AI 서비스 신뢰성을 위한 영향평가로서 2단계의 안정성 영향평가 항목에 추가하여 윤리, 이용자 권익 등의 반영 여부를 평가함으로써 완성된 버전의 영향평가 기준을 개발하는 것이다.

또한, 인공지능에 따른 사회의 변화 양상을 파악하기 위해 정책 현황 분석 및 우리나라 수준 조사, 국가사회 전반의 영향·성과 분석, 분야별 영향평가로 세분화하여 추진하는 것도 고려가 필요하다.

V. 결론 및 논의

본 연구의 목적은 영향평가에 대한 정의부터 국내외 인공지능 서비스 영향평가 사례에 대한 심도있는 분석을

바탕으로 인공지능 서비스 영향평가 추진 방향에 대한 정책적 제언을 수립하는 것이다. 이를 위해서 국내외 타 분야에서 추진된 영향평가 사례, 인공지능 서비스 특성을 고려한 국내외 영향평가 사례 및 추진시 고려 요소를 분석하였다.

분석 결과를 살펴보면 다음과 같이 정리하였다.

1단계는 AI 인덱스 확장 기반 영향평가로서 기존의 연구 된 내용을 기반으로 국내에 적합한 방안 연구를 통해 구체적으로 시장에 적용하여 평가할 수 있는 대안을 마련하는 것이다. 또한, AI 영향평가는 정책 또는 사업의 허가·인가·승인·면허 등 행정행위를 위한 것이 아닌, 정책의 합리적 개선을 유도하기 위한다는 취지를 고려할 때 ‘규범형’이 바람직하다.

2단계는 AI 서비스 안정성을 위한 영향평가로서 인공지능 서비스의 설계/개발 단계에서 사이버 레질리언스 역량과 프라이버시 보호 기능을 평가하는 것으로, 적용 대상은 공공기관에서 수행하는 인공지능 서비스를 대상으로 하되 민간기업의 경우, 인센티브를 제공하는 방식으로 간접적인 유도방안을 제시한다. 3단계는 AI 서비스 신뢰성을 위한 영향평가로서 2단계의 안정성 영향평가 항목에 추가하여 윤리, 이용자 권익 등의 반영 여부를 평가함으로써 완성된 버전의 영향평가 기준을 개발하는 것이다.

또한, 인공지능에 따른 사회의 변화 양상을 파악하기 위해 정책 현황 분석 및 우리나라 수준 조사, 국가사회 전반의 영향·성과 분석, 분야별 영향평가로 세분화하여 추진하는 것도 고려가 필요하다.

국내 인공지능 서비스 영향평가는 다소 광범위하고 포괄적이며, 시점도 사전적 예방 수단에 그치지 않고 상시적·사후적 위험성 관리를 예정하고 있다. 이 인공지능 서비스 영향평가 결과가 정책에 반영하는 환류 체계까지 포함된다면 국내의 인공지능 경쟁력 강화에 도움이 되는 정책 수단이 될 수 있다.

본 연구의 한계는 기존 연구된 문헌에 대한 분석으로 향후에는 정책 설계자, 시민단체 등 다양한 참여자를 통한 조사 및 분석이 필요하다.

References

- [1] Ministry of Science and ICT, "NATIONAL STRATEGY FOR AI", 2019
- [2] Ministry of Science and ICT, "Year Book of information

- Society Statistics", 2021
- [3] McKinsey, The state of AI in 2021
<https://www.mckinsey.com/business-functions/quantumblack/our-insights/global-survey-the-state-of-ai-in-2021>
- [4] European Commission, "Regulatory framework proposal on artificial intelligence", 2021,
<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>
- [5] UC Berkeley, Guidance for the Development of AI Risk and Impact Assessments, 2021,
<https://cltc.berkeley.edu/2021/08/09/guidance-for-the-development-of-ai-risk-and-impact-assessments/>
- [6] Kim jinyeoul. "A study on Policy Impact Assessment and customer impact assessment", Customer policy, no. 109, 1-25, 2020
- [7] Chang, Inho. "A Study on the Impact Assessment System for the Realization of Sustainability", Study on The American Constitution Institution of American Constitution, Vol.24 No.2 Aug. 2013
- [8] Seo Jiyoung. "Technological Impact Assessment Improvement Plan for Responsible Research and Innovation" STEPI Insight, no. 157, 1-33. 2015
- [9] UNESCO, Recommendation on the Ethics of AI, 2021
- [10] Alessandro Mantelero, AI AND BIG DATA: A BLUEPRINT FOR A HUMAN RIGHTS, SOCIAL AND ETHICAL IMPACT ASSESSMENT, Computer Law & Security Review, Vol 34, No4, 2018.
DOI:<https://doi.org/10.1016/j.clsr.2018.05.017>
- [11] OECD. Improving Policy Instruments through Impact Assessment, SIGMA Papers, No. 31. OECD, 2001
- [12] Stanford Univ. HAI, AI index report, 2021.
- [13] NIST, AI RISK MANAGEMENT FRAMEWORK, 2022,
<https://www.nist.gov/itl/ai-risk-management-framework>
- [14] EU, Sherpa project: <https://www.project-sherpa.eu/>
- [15] EU, AI Impact Assessment, Project Sherpa, 2021
- [16] UK, Algorithmic impact assessment in healthcare in NHS, 2022,
<https://www.adalovelaceinstitute.org/project/algorithmic-impact-assessment-healthcare/>
- [17] Kazim, Emre, Danielle Mendes Thame Denny, and Adriano Koshiyama. "Ai Auditing and Impact Assessment: According to the Uk Information Commissioner's Office." AI and Ethics 1, no. 3, 301-10. 2021.
DOI: <https://doi.org/10.1007/s43681-021-00039-2>
- [18] Kim Myung-ju. "The necessity of artificial intelligence ethics and domestic and international trends." THE JOURNAL OF KOREAN INSTITUTE OF COMMUNICATIONS AND INFORMATION SCIENCES, 34, no. 10, 45-54, 2017
- [19] ECP, AI Impact Assessment(AIIA), 2019
- [20] Government of Canada, Algorithmic Impact Assessment,<https://canada-ca.github.io/aia-eia-js/>
- [21] Schiff, Daniel, Aladdin Ayeshe, Laura Musikanski, and John C Havens. Ieee 7010: A New Standard for Assessing the Well-Being Implications of Artificial Intelligence. 2020 IEEE international conference on systems, man, and cybernetics (SMC): IEEE, 2020. DOI:<https://doi.org/10.1109/smc42975.2020.9283454>
- [22] Yoo, Soonduck. "A Study on the Impact and Activation Plan of Unmanned Aerial Vehicle Service." The Journal of the Institute of Internet, Broadcasting and Communication 22, 2022.
DOI:<https://doi.org/10.7236/JIIBC.2022.22.2.1>
- [23] Jeong-Keun Yun. "A Study on the Policy Innovation Plan for Public Technology Commercialization." Journal of the Korea Academia-Industrial cooperation Society, 22, 2021.
DOI:<https://doi.org/10.5762/KAIS.2021.22.2.212>

저 자 소 개

신 선 영(정회원)



- 동국대학교 컴퓨터공학과
- 연세대학교 산업공학 석사
- 경북대학교 경영정보학 박사
- 2001 ~ 현재 : 한국지능정보사회진흥원(NIA) AI 미래전략센터장
- 관심분야 : 데이터 기반 의사결정, 빅데이터 분석, AI 사회적 영향평가, ICT 정책 수립